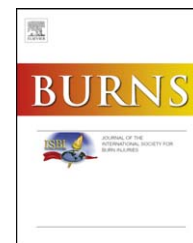


available at www.sciencedirect.comjournal homepage: www.elsevier.com/locate/burns

Reliability of subjective wound assessment

M.C.T. Bloemen^{a,b}, P.P.M. van Zuijlen^{a,b,c,d,e}, E. Middelkoop^{a,b,c,*}

^a Association of Dutch Burn Centres, Beverwijk, The Netherlands

^b Burn Centre, Red Cross Hospital, Beverwijk, The Netherlands

^c Department of Plastic, Reconstructive and Hand Surgery, VU University Medical Centre, Amsterdam, The Netherlands

^d Department of Plastic, Reconstructive and Hand Surgery, Red Cross Hospital, Beverwijk, The Netherlands

^e Department of Plastic, Reconstructive and Hand Surgery, Academic Medical Centre, Amsterdam, The Netherlands

ARTICLE INFO

Article history:

Accepted 2 February 2011

Keywords:

Wound

Assessment

Reliability

Subjective

Graft take

Epithelialisation

Experience

ABSTRACT

Introduction: Assessment of the take of split-skin graft and the rate of epithelialisation are important parameters in burn surgery. Such parameters are normally estimated by the clinician in a bedside procedure. This study investigates whether this subjective assessment is reliable for graft take and wound epithelialisation.

Methods: Observers involved in the field of burns (experienced, medium-experienced and inexperienced observers), and dermatologists specialized in the field of wound healing evaluated the percentage graft take and epithelialisation in 50 photographic skin-grafted burn wounds. Reliability was tested using the intraclass correlation coefficient (ICC).

Results: Intra- and interobserver reliability of parameter graft take was highest within the experienced observers (ICC average > 0.91), followed by medium- and inexperienced observers (ICC average > 0.80 and ICC average > 0.68). Parameter epithelialisation showed the same pattern of intra- and interobserver ICC scores (experienced > medium > inexperienced). Interobserver ICC single scores of the experienced group were reasonable to good. Interobserver reliability of the dermatologists was similar to medium-experienced observers.

Conclusions: Our data show that one experienced observer can obtain adequate reliable results by means of a single assessment of graft take and epithelialisation. Furthermore, experience of the observer results in an increase of reliability.

© 2011 Elsevier Ltd and ISBI. All rights reserved.

1. Introduction

The quantification of parameters of wound healing remains a frequent topic of discussion amongst dermatologists, surgeons, plastic surgeons, and clinimetricians. Assessment of these parameters allows clinicians and researchers to evaluate the progress of wound healing. Additionally, the effectiveness of treatments to improve wound management, can be measured. Some of the relevant parameters for the evaluation of wound healing are wound area, epithelialisation, and take of a split-skin graft, i.e. graft take. Significant research is

performed on wound surface area measurement, however, a paucity exists on the assessment of epithelialisation and graft take. Nevertheless, especially these parameters are most valuable for the assessment of wounds that received meshed autologous split-skin grafts.

To our knowledge, no studies have been reported on objective assessment of graft take. The objective assessment of epithelialisation has been investigated in some studies, for instance the measurement of re-epithelialisation with photoplanimetry [1], optical coherence tomography [2], or measuring water evaporation [3], however, these techniques were not

* Corresponding author at: Association of Dutch Burn Centres, Postbus 1015, 1940 EA Beverwijk, The Netherlands. Tel.: +31 251 275500; fax: +31 251 216059.

E-mail addresses: e.middelkoop@vumc.nl, emiddelkoop@rkz.nl (E. Middelkoop).

0305-4179/\$36.00 © 2011 Elsevier Ltd and ISBI. All rights reserved.

doi:10.1016/j.burns.2011.02.004

evaluated sufficiently on the clinimetric properties reliability, validity and feasibility. Epithelialisation has also been evaluated histologically by means of a biopsy [1,4–6], however this method is a burden to the patient and additionally, the biopsy creates a new wound resulting in additional pain, the risk of infection, and scarring. Objective tools which are feasible, appear to be scarce and therefore, the most important and optimal evaluation of the parameters graft take and epithelialisation still seems to be the clinician's subjective assessment.

In the majority of published studies, graft take and epithelialisation were carried out subjectively, either by means of a bedside assessment or by means of viewing photographic images [5–9]. The subjective evaluation is fast and easy to perform, however, it could be inconsistent. In addition, wounds are often evaluated by different observers with diverse opinions and different experience, which will inevitably introduce extra bias. Especially in multicentre clinical trials, it is desirable for the outcome of the trial that observers assess the wounds in the same way. Nevertheless, it seems that the subjective assessment has not been investigated thoroughly; only one report examined the subjective evaluation of the parameter epithelialisation [1]. Therefore, this study was set up to investigate the intra- and interobserver reliability of the parameters graft take and epithelialisation in transplanted wounds. A second aim of this study was to examine the relation between clinical experience and the reliability of the assessment. For this reason, the evaluation was performed by observers with different levels of expertise.

2. Methods

2.1. Study design

Eight clinicians, experienced in the field of burn surgery at the Burn Centres of the Netherlands, and four dermatologists, active in the field of wound healing at three different University Medical Centres of the Netherlands were asked to evaluate 50 photographic images of 50 transplanted burn wounds. In addition, four medium-experienced clinicians and three inexperienced students of the Dutch Burn Centres evaluated the photographic wounds. Experienced observers had more than 10 years practice in burns, medium-experienced observers had worked less than three years in burn surgery or were not trained as (plastic) surgeons, and the inexperienced students had no clinical experience in wound healing or burns. Besides the assessment of burn clinicians, dermatologists were also selected for this study in order to examine reliability of wound assessment broadly, i.e. in more specialties.

The photographic wounds assessed in this study were collected from burn patients consecutively treated in an ongoing multicenter randomised controlled trial (RCT). The trial, performed in the three Dutch Burn Centres, investigates the effectiveness of dermal substitution in combination with topical negative pressure in the treatment of acute burns that require skin grafting. The study protocol was approved by the medical ethics committee (M07-035) and registered at ClinicalTrials.gov (ID NCT00548314). Patients (with informed consent) were allocated to the following four treatment groups: (1) dermal substitute in combination with

a split-skin graft (SSG) and topical negative pressure; (2) dermal substitute in combination with a SSG; (3) SSG and topical negative pressure; (4) SSG alone. All skin grafts were expanded with an 1:1.5 ratio. Four to seven days after skin transplantation, all wounds were photographed using a commercially available digital camera, as well as assessed in a bedside procedure by an independent clinician. In addition, progress of wound healing was followed. The digital photographs of the first 50 patients of the above mentioned clinical trial were selected for the evaluation by the observers of the present study. Fig. 1 shows an overview of the study design.

Each observer independently examined and documented the percentage of graft take and the percentage of wound epithelialisation. The percentage of graft take was defined as the percentage of the skin graft on the wound which appeared vital and adherent. The percentage of wound epithelialisation was defined as the percentage of the grafted wound which was closed. The total grafted wound was evaluated; this also included wound areas with graft loss and graft displacement. Observers involved in the field of burns were asked to assess the photographs twice, with an interval of 4–6 weeks. To avoid a memory bias, the time interval should not be too short, as observers may remember their first response. The appropriate interval is reported to vary from 1 h to a year depending on the task, but generally speaking 2–14 days is usual [10]. For logistical reasons, the dermatologists only evaluated the wounds once, therefore only the interobserver reliability could be established.

Digital photographs were all in focus, with a complete view of the relevant wound area and a high enough resolution to be able to judge wound aspects in detail. Accordingly, photographs were randomly arranged in a PowerPoint presentation. Neither wound treatment nor any patient information could be derived from the sequence of the photographs. Preceding the first assessment of the photographs, the observers received a personal instruction in which the definitions of the wound parameters were clarified. Preceding the second assessment of the photographs, the observers were given the choice to receive the personal instruction again or to follow the digital instruction within the PowerPoint presentation. The observers were not informed about their previous assessment scores.

2.2. Statistical analysis

Data were analysed with the statistical program SPSS for Windows 18.0 (SPSS Inc., Chicago, USA). In clinimetric research, a number of 50 measurements or more is considered sufficient. The intraobserver reliability of photographic wound assessment measures the agreement of one observer when assessing the same image. The interobserver reliability measures the agreement between several observers when assessing the same image. Analysis was performed by means of the intraclass correlation. Also the 95% confidence interval was calculated. The two-way random-effect model and the absolute agreement type for a single and average measurement were selected for the calculations of the intraclass correlation [11]. This calculation provides a single measure intraclass correlation coefficient (ICC), which is based on a

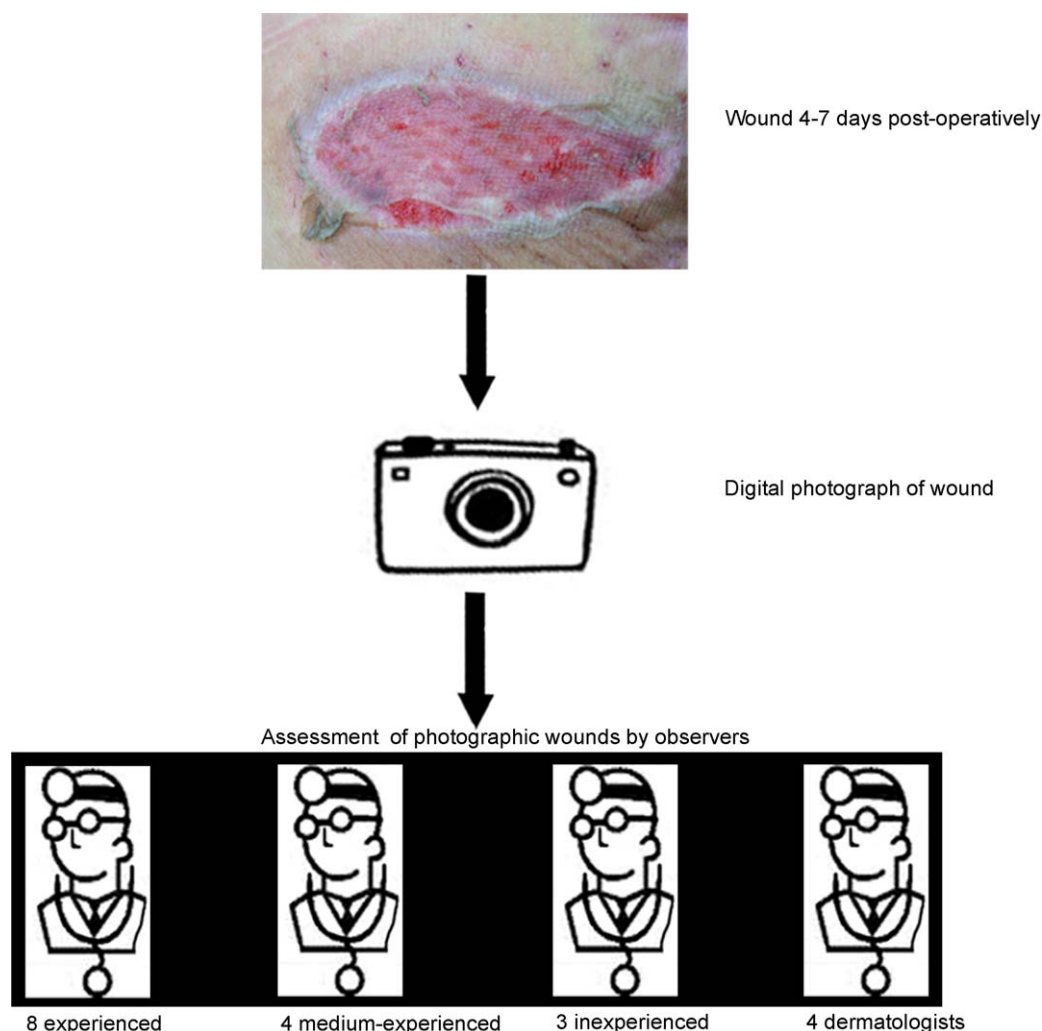


Fig. 1 – Study design.

single measurement and an average measure ICC which is based on the average measurements of all observers. According to Fleiss, a coefficient below 0.4 represents poor reliability, values above 0.75 represent excellent reliability and values between 0.4 and 0.75 stand for fair to good reliability [12]. The standard error of measurement (SE_{meas}) was used for the calculation of errors between measurements. The coefficient of variation (CV) was calculated using the following formula: $CV = SE_{meas}/\text{mean} \times 100$. A low CV percentage stands for less variation in error than a high CV.

3. Results

The first 50 patients of the above mentioned RCT were treated in the Burns Centres of Beverwijk and Rotterdam in the Netherlands and were used for this clinimetric study. The mean day that wounds were evaluated was 4.8 ± 0.6 days post-operatively. Forty-five out of 50 wounds healed without regrafting. Table 1 shows the mean values of parameters graft take and epithelialisation of the different observer groups.

3.1. Intraobserver reliability

The intraobserver ICC scores calculated for the percentage graft take and epithelialisation were highest in the group of the experienced observers (Table 2). This was followed by the intraobserver ICC scores of the medium-experienced observers. The ICC scores were lowest in the group with the inexperienced observers. The mean CV of the both parameters was lowest in the experienced group and highest in the inexperienced observers.

Table 1 – Mean percentages of graft take and epithelialisation.

Observer	Graft take (%)	Epithelialisation (%)
Experienced observers (n = 8)	84.0 ± 17.9	72.1 ± 16.7
Medium-experienced observers (n = 4)	83.8 ± 19.0	73.9 ± 18.8
Inexperienced observers (n = 3)	70.5 ± 16.0	62.7 ± 13.9
Dermatologists (n = 4)	77.0 ± 17.6	63.3 ± 18.4

Table 2 – Intraobserver reliability of the different observers.

	Mean ICC Experienced	Mean ICC Medium-experienced	Mean ICC Inexperienced
Graft take			
Single measure	0.84	0.80	0.58
Average measure	0.91	0.89	0.72
CV (%)	9.6	11.6	19.3
Epithelialisation			
Single measure	0.79	0.79	0.48
Average measure	0.88	0.88	0.63
CV (%)	12.7	13.2	22.6
The intraclass correlation coefficient (ICC) of a single and average measure of both wound parameters. The coefficient of variation (CV = $SE_{meas}/mean \times 100$) is given.			

Table 3 – Interobserver reliability of the different observers.

	ICC (95% CI) Experienced	ICC (95% CI) Medium-experienced	ICC (95% CI) Inexperienced	ICC (95% CI) Dermatologists
Graft take				
Single measure	0.66 (0.55–0.76)	0.49 (0.33–0.65)	0.42 (0.25–0.58)	0.52 (0.38–0.66)
Average measure	0.94 (0.91–0.96)	0.80 (0.67–0.88)	0.68 (0.49–0.81)	0.81 (0.71–0.89)
CV (%)	13.6	18.2	23.9	21.9
Epithelialisation				
Single measure	0.56 (0.43–0.69)	0.54 (0.40–0.67)	0.41 (0.24–0.58)	0.53 (0.37–0.67)
Average measure	0.91 (0.86–0.95)	0.82 (0.73–0.89)	0.68 (0.48–0.81)	0.82 (0.70–0.89)
CV (%)	17.1	19.6	25.5	25.5
The intraclass correlation coefficient (ICC) with 95% confidence interval (CI) of a single and average measure of both wound parameters. The coefficient of variation (CV = $SE_{meas}/mean \times 100$) is given.				

3.2. Interobserver reliability

The same pattern was seen for the interobserver reliability of the different observer groups (Table 3): interobserver ICC scores were highest in the group with the experienced observers and lowest in the group with the inexperienced observers. ICC scores of the dermatologists were equal to the medium-experienced observers. Least variation (CV) was seen in the measurements of the experienced observers.

3.3. Correlation

To correlate the outcome of photographic assessments versus bedside evaluation, we compared the photographic wound assessment of the experienced observers to the bedside assessment that was carried out by an independent clinician, simultaneously with the photography of the wound. The correlation of these two assessments for the parameter graft take was shown to be good ($r = 0.69$, $p < 0.001$, Pearson's correlation). The bedside assessment of the parameter epithelialisation also showed a good correlation with the photographic assessment of the experienced observers ($r = 0.71$, $p < 0.001$, Pearson's correlation).

the medium- and inexperienced observers. The intra- and interobserver ICC scores of the inexperienced observers were lowest of all observers, but still reasonably good (ICC average > 0.63). First of all, these high intraobserver scores of the experienced or medium-experienced observers demonstrate that in the evaluation of graft take and epithelialisation rate, only one photographic assessment is needed. The interobserver single measure scores show that only one experienced observer is needed for fair to good reliability of the evaluation of these wound parameters. As the single measure ICC score was below 0.75, we recommend performing evaluation of graft take and epithelialisation by two experienced observers for a higher reliability.

Comparable with our results, Laplaud et al. previously reported a high intra- and interobserver reliability for subjective wound assessment [13]. However, in their study the percentage of fibrin was assessed instead of the percentage of epithelialisation. In another study, the correlation of the subjective evaluation of epithelialisation of two observers was reported to be reasonable, although no statistical analysis was shown [1]. Vermeulen et al. investigated the agreement of observers assessing wound healing using a classification scheme based on colours [14]. They found a moderate to good agreement between observers on wound colour in open surgical wounds.

Second, our results demonstrate that reliability of subjective wound assessment increases with experience of the observers. The experience of observers is probably derived from the knowledge of the various wound aspects, such as re-epithelialisation, necrosis, or the vitality of the graft. In comparison with the ICC scores, the variances were highest in the inexperienced group. These findings are in agreement with

4. Discussion

In this study, the reliability of subjective assessment of graft take and epithelialisation was clinimetrically investigated for the first time. Observers experienced in burns, obtained intra- and interobserver average measure scores that are excellent (ICC > 0.88). Their ICC scores were higher than the scores of

our expectations as experience is thought to give a more consistent assessment. The data can be used for comparison with data of future research in this area. The need for experience in wound assessment has been reported previously [15]. However, in one report, the opposite was stated in the estimation of wound surface percentages; no relationship was found between the clinical experience of the observer and the accuracy of the visual estimation [16]. In our study, the group of dermatologists showed a good interobserver reliability, despite the fact that these clinicians had little experience in burns. Their knowledge of wound healing in chronic wounds is likely the reason for the relatively high reliability scores. These data emphasize the value of expertise in wound healing for a reliable assessment.

As previously mentioned, the ICC average measure scores were excellent, however an improvement of the single measure scores would be desirable. The single measure score could have been higher using more standardized study methods. First of all, we investigated the reliability of wound assessment by means of photographic images. However, the most ideal study design would have been to test the reliability during a bedside procedure. In this study, it was logistically not possible to arrange that all 19 observers assessed the wounds of 50 patients during a bedside procedure. As it was previously demonstrated that wound assessment at bedside and visual examination of photographs are equivalent, this research was based on a series of photographic wounds [13,15,17,18]. In this study, good results were obtained with the use of photographic images, therefore, the subjective assessment performed during a bedside procedure is likely to show a higher reliability.

Second, different treatment methods were used for the wounds in the clinical trial. This could have had an effect on the observer's wound assessment. For instance, it may be more difficult to evaluate wounds treated with a split-skin graft in combination with a dermal substitute. In an additional analysis of the experienced observers, we investigated reliability in separate groups: wounds treated with and without the dermal substitute. As expected, interobserver reliability scores were higher when only the scores of wounds treated without the substitute were analysed compared to the scores of the total group (data not shown). The substitute may cause a delayed vascular ingrowth of the graft and as a consequence, the graft could look less vital (e.g. other colour), several days post-operatively, compared to the aspect of the skin graft that is placed directly onto the wound. Most likely, the ICC scores would have been higher if all wounds were treated with a split-skin graft alone. Finally, the following two study aspects could also improve the intra- and interobserver reliability: (1) the conditions under which the photographic images were made, which were not fixed in this study (e.g. light, camera-wound distance, camera type) and (2) the relationship or collaboration of the observers, since in this study the observers operated in different teams and centres. However, we chose to perform this study with a realistic and clinically relevant wound population.

Summarizing, for the first time subjective assessment of graft take and epithelialisation was clinimetrically investigated and as a result we can make the following recommendations:

- Experience is highly relevant for a reliable assessment.
- Assessment of one observer (with experience) has an adequate reliability and a second observer will increase reliability.
- One (photographic) measurement per observer is sufficient.
- Reliability is likely to increase in a more standardized setting (e.g. bedside evaluation by the same clinician and similar wound treatment in all wounds).

Conflict of interest statement

None.

Acknowledgements

The authors would like to thank all observers for their contribution to the data: G. Beerthuijzen, H. Boxma, E. de Boer, J. Dokter, R. Kreis, K. Munte, I. Oen, E. Res, J. Schalkwijk, S. Scholten, V. Sigurdsson, F. Tempelman, M. van der Wal, M. van Leeuwen, C. van Montfrans, N. van Vucht, P. Verhaegen, J. Vloemans. This study was partially financed by the Dutch Burns Foundation (grant number 07.109), which had no involvement in the study.

REFERENCES

- [1] Hauser J, Lehnhardt M, Daigeler A, Langer S, Steinau HU, Vogt PM. Photoplanimetric evaluation and impedance measurement of split-thickness skin grafts: a new model for objective wound-healing assessment in clinical trials. *Skin Res Technol* 2009;15(2):168-71.
- [2] Singer AJ, Wang Z, McClain SA, Pan Y. Optical coherence tomography: a noninvasive method to assess wound reepithelialization. *Acad Emerg Med* 2007;14(5):387-91.
- [3] Surinchak JS, Malinowski JA, Wilson DR, Maibach HI. Skin wound healing determined by water loss. *J Surg Res* 1985;38(3):258-62.
- [4] Kaawach WF, Oliver AM, Weiler-Mithoff E, Abramovich DR, Rayner CR. Survival assessment of cultured epidermal allografts applied onto partial-thickness burn wounds. *Br J Plast Surg* 1991;44(5):321-4.
- [5] Danielsen P, Jorgensen B, Karlsmark T, Jorgensen LN, Agren MS. Effect of topical autologous platelet-rich fibrin versus no intervention on epithelialization of donor sites and meshed split-thickness skin autografts: a randomized clinical trial. *Plast Reconstr Surg* 2008;122(5):1431-40.
- [6] Stone CA, Wright H, Clarke T, Powell R, Devaraj VS. Healing at skin graft donor sites dressed with chitosan. *Br J Plast Surg* 2000;53(7):601-6.
- [7] Silver GM, Robertson SW, Halerz MM, Conrad P, Supple KG, Gamelli RL. A silver-coated antimicrobial barrier dressing used postoperatively on meshed autografts: a dressing comparison study. *J Burn Care Res* 2007;28(5):715-9.
- [8] van Zuijlen PP, van Trier AJ, Vloemans JF, Groenevelt F, Kreis RW, Middelkoop E. Graft survival and effectiveness of dermal substitution in burns and reconstructive surgery in a one-stage grafting model. *Plast Reconstr Surg* 2000;106(3):615-23.
- [9] Innes ME, Umraw N, Fish JS, Gomez M, Cartotto RC. The use of silver coated dressings on donor site wounds: a

- prospective, controlled matched pair study. *Burns* 2001;27(6):621-7.
- [10] Streiner DL, Norman GR. *Health Measurement Scales: a practical guide to their development and use: reliability*, 4th ed., Oxford: Oxford University Press; 2008.
- [11] Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing reliability. *Psychol Bull* 1979;2:420-8.
- [12] Fleiss JL. *Statistical methods for rates and proportions*. New York: Wiley; 1981.
- [13] Laplaud AL, Blaizot X, Gaillard C, Morice A, Lebreuilly I, Clement C, et al. Wound debridement: comparative reliability of three methods for measuring fibrin percentage in chronic wounds. *Wound Repair Regen* 2010;18(1):13-20.
- [14] Vermeulen H, Ubbink DT, Schreuder SM, Lubbers MJ. Inter- and intra-observer (dis)agreement among nurses and doctors to classify colour and exudation of open surgical wounds according to the Red-Yellow-Black scheme. *J Clin Nurs* 2007;16(7):1270-7.
- [15] Houghton PE, Kincaid CB, Campbell KE, Woodbury MG, Keast DH. Photographic assessment of the appearance of chronic pressure and leg ulcers. *Ostomy Wound Manage* 2000;46(4):20-6. 8-30.
- [16] Mekkes JR, Westerhof W. Image processing in the study of wound healing. *Clin Dermatol* 1995;13(4):401-7.
- [17] Wirthlin DJ, Buradagunta S, Edwards RA, Brewster DC, Cambria RP, Gertler JP, et al. Telemedicine in vascular surgery: feasibility of digital imaging for remote management of wounds. *J Vasc Surg* 1998;27(6):1089-99. discussion 99-100.
- [18] Murphy Jr RX, Bain MA, Wasser TE, Wilson E, Okunski WJ. The reliability of digital imaging in the remote assessment of wounds: defining a standard. *Ann Plast Surg* 2006;56(4):431-6.